

# Bias in the analysis of multilingual legislative speech

Laura Hollink, Astrid van Aggelen, Jacco van Ossenbruggen  
*Centrum Wiskunde & Informatica, Amsterdam, The Netherlands*  
l.hollink@cwi.nl

In this paper we investigate the application of natural language processing tools to the multilingual proceedings of the European Parliament. This work is part of a study in which we explore (1) how subcorpora in different languages may lead to different conclusions about the political landscape, (2) how to determine what a potential language-related bias originates from, and (3) to what extent we can limit or even prevent an unwanted language-bias.

Parliamentary speech has been used to study party positions [1,2,3], issue selection [4,5,6,7] and the level of disagreement within a debate [8]. Many studies have moved away from manual coding (which is done in e.g. [4,5]) and instead position speech texts on one or more (latent) dimensions in statistical models based on relative word frequencies [1,2,3,6,7,8], often in combination with basic pre-processing steps such as stemming and stopping. These models and tools, while imperative to analyse bigger datasets, add a source of errors and bias. One source of potential bias comes from the fact that the used tools perform differently on different languages. Considering that the aforementioned studies were carried out on the European, Irish, US, Spanish, Norwegian and Swedish legislatures, the comparability and reproducibility of the results for different languages is unclear.

In the European Parliament, the spoken accounts appear in (currently) 24 languages. Here, the uncertainty stems not only from tools that perform differently on each language, but also from the fact that the availability of data in each language varies. Members of Parliament (MEPs) are free to speak in any of the official languages. Speeches are sometimes translated into (some) other languages, depending on prioritization with the EP, specific translation-requests of the members and (supposedly) budgetary constraints. Thus, we are left with 24 subcorpora of varying size, one per language, including both original and translated speech.

The need to study language-effects in this context has been recognised before. Proksch *et al.* [3] reported a modest language-effect<sup>1</sup> in their study of party positions in the European Parliament, which they ascribed to translation rather than actual differences in position taking between three countries. However, while the overall effect may be small, we argue that specific local effects could still lead to significant biases in the results. For example, French translations of German texts seemed to systematically get a more neutral position than the original text, while the opposite was not the case. It is important to realise that the proceedings of the European Parliament are not only a corpus for researchers. Residents of the European Union have a right to access these documents in order to make informed votes and to hold the MEPs accountable<sup>2</sup>. This right would be compromised when French speaking citizens come to different conclusions about what has been discussed than German speaking citizens. Our aim is to gain insight into how working with subcorpora in different languages may lead to different conclusions about the political landscape.

---

<sup>1</sup> A correlation coefficient ranging between 0.86 and 0.93 when comparing party positions derived from texts in German, French and English [3].

<sup>2</sup> Regulation (EC) No 1049/2001 of the European Parliament and of the Council

In this study, we use the data provided by the Talk of Europe project [9], in which speech transcripts and all available translations were crawled from the website of the EP<sup>3</sup>, and translated into the semantic web format RDF. Data is available from 1999 to 2015 and contains around 300K speeches in 22K debates. We apply topic detection to six language-specific subcorpora of the proceedings of the European Parliament: German, English, French, Italian, Spanish and Dutch. We use the JEX software developed by the European Commission's Joint Research Centre, which learns multi-label categorisation rules from documents that were previously manually indexed using the multilingual Eurovoc thesaurus [10]. The advantage of using this tool over, for instance, widely used topic modeling approaches such as LDA [11], is that the output is directly comparable across languages: the tool uses a single thesaurus, Eurovoc, to classify documents in each language, and concepts in the Eurovoc thesaurus have labels in all languages. In a later stage of the study, we plan to include other topic detection techniques, and widen the scope to all EU languages.

Over 2000 distinct Eurovoc topics were detected in the six subcorpora. The frequency distributions over topics vary per language. Figure 1 visualises the distance between languages. We use Kullback–Leibler divergence [12], a non-symmetric measure for the difference between two distributions. A higher score, visualized as a redder colour, signifies a greater distance. For example, Italian and French are relatively close, while Spanish and German are far apart. There are four hypotheses as to what these differences originate from:

1. MEPs speaking one language indeed speak about different topics than their colleagues who speak in another language.
2. There is a bias in the selection of speeches that are being translated.
3. There is a bias in how certain topics are translated, e.g. translators use more ambiguous or polarized language.
4. The topic detection tool works differently on one language than on another.

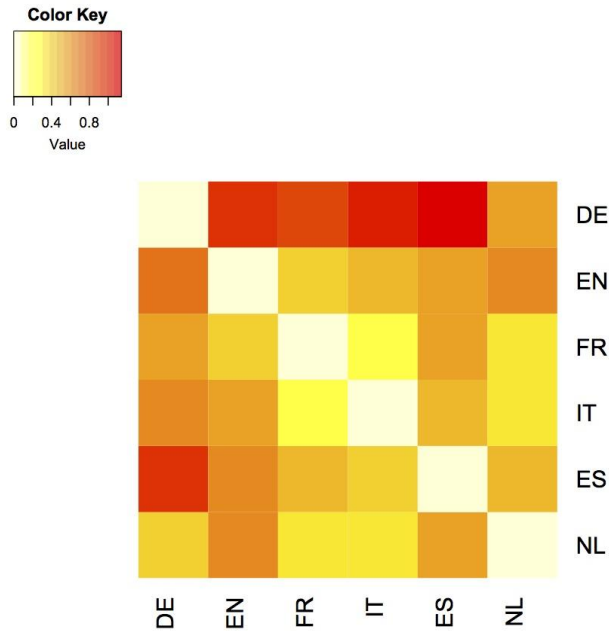


Figure 1: Heatmap of differences between topic distributions in languages.

In our presentation, we will tackle this issue from two sides. Firstly, we compare different subsets of topics based on whether or not speeches were translated, and to which languages, to explore hypotheses 1 and 2. Then, to study hypothesis 4 (and to a lesser extent hypothesis 3) we zoom into topics that appear to be

<sup>3</sup> <http://www.europarl.europa.eu>

particularly distinctive between languages, and compare the topic annotations to what was actually said in the debates. As an example of the latter method, Figure 2 shows the differences in frequency of the detected topics “nuclear weapons” and “nuclear energy”. Remarkably, only French and Italian speeches seem to be about nuclear weapons, while English and Spanish speeches are often about nuclear energy. As a comparison, Figure 3 plots the occurrences of the phrases “nuclear weapons” and “nuclear energy” (and translations thereof) in the raw speech texts. Here, part of the effect is gone, suggesting an error of the topic annotation software, while part of the effect remains - German texts indeed seem to talk less about both nuclear weapons and nuclear energy.

With this study, we aim to contribute to the discussion about systematic methods for tool criticism and source criticism in a complex multilingual context like the European Parliament.

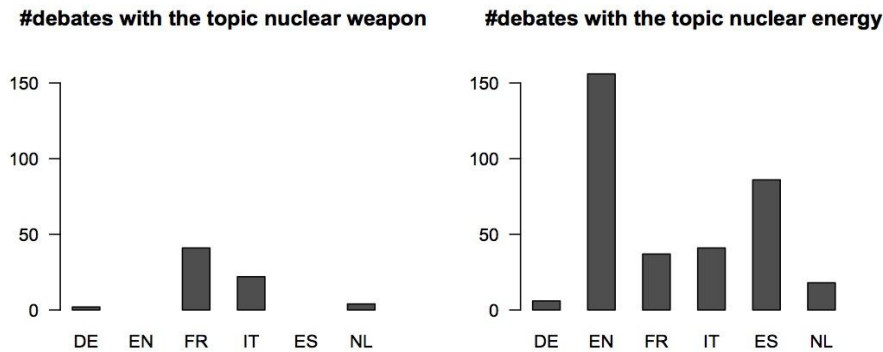


Figure 2: Frequency of topics in debates.

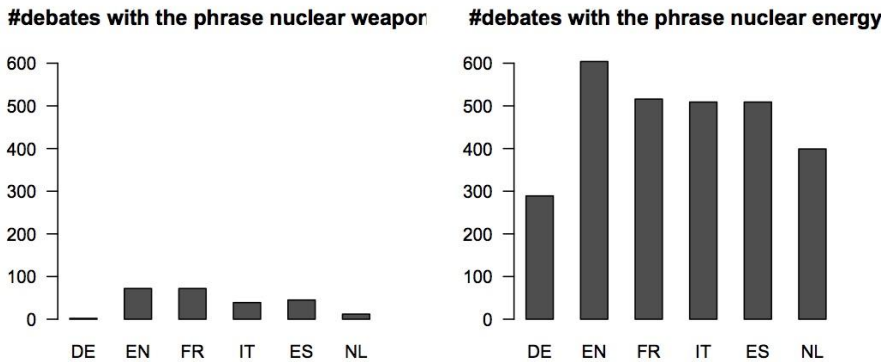


Figure 3: Frequency of phrases in debate texts.

## References

- [1] Benoit, Kenneth, and Michael Laver Nd. Estimating Irish Party Positions Using Computer Wordscoring: The 2002 Elections. *Irish Political Studies* Vol. 18 , Iss. 1, 2003.
- [2] Laver, Michael J., Kenneth R. Benoit, and John Garry. Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review* 97(2):311–31, 2003.
- [3] Proksch, S.-O. and Slapin, J.B. Position Taking in European Parliament Speeches, *British Journal of Political Science*, 40(3), pp. 587–611, 2010.

- [4] Hanna Bäck, Marc Debus & Jochen Müller. Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag. *Political Research Quarterly* 67: 504–518, 2014.
- [5] Markus Baumann. Constituency Demands and Limited Supplies: Comparing Personal Issue Emphases in Co-sponsorship of Bills and Legislative Speech. *Scandinavian Political Studies*, Vol. 39, issue 4, pp. 366-387, 2016.
- [6] Pardos-Prado, Sergi, and Iñaki Sagarzazu. The Political Conditioning of Subjective Economic Evaluations: The Role of Party Discourse. *British Journal of Political Science* 46(4), 799-823, 2016.
- [7] Kevin M. Quinn , Burt L. Monroe , Michael Colaresi , Michael H. Crespin , Dragomir R. Radev. An automated method of topic-coding legislative speech over time with application to the 105th-108th US Senate. *Midwest Political Science Association Meeting*. 2006.
- [8] Benjamin E. Lauderdale, Alexander Herzog. Measuring Political Positions from Legislative Speech. *Polit Anal*; 24 (3): 374-394, 2016.
- [9] Astrid van Aggelen, Laura Hollink, Max Kemman, Martijn Kleppe, and Henri Beunders. The debates of the european parliament as linked open data. *Semantic Web*, 8(2):271–281, 2017.
- [10] Pouliquen Bruno, Steinberger Ralf, Camelia Ignat. Automatic Annotation of Multilingual Text Collections with a Conceptual Thesaurus. In *Proceedings of the Workshop Ontologies and Information Extraction at the Summer School The Semantic Web and Language Technology - Its Potential and Practicalities (EUROLAN'2003)*. Bucharest, Romania, 28 July - 8 August 2003.
- [11] Blei, David M., Ng, Andrew Y., Jordan, Michael I. Lafferty, John, ed. Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 3 (4–5): pp. 993–1022, 2003.
- [12] Kullback, S., Leibler, R.A. On information and sufficiency. *Annals of Mathematical Statistics*. 22 (1): 79–86, 1951.